

서울시 아동보호전문기관 우선 입지 선정



Contents



01

주제 및 선정 배경

02

전처리 및 EDA

03

회귀 모델링

04

다음 주 예고



01

주제 및 선정 배경

1) 선정 배경

2) 주제 소개

01. 선정 배경 주제 소개

02. 전처리 및 EDA

03. 회귀 모델링

04. 다음주 예고

01. 주제 선정 배경

YTN | 2015.05.21. | 네이버뉴스
'칠곡 계모' 항소심 징역 15년
대구고등법원은 8살 난 의붓딸을 학대하고 폭행해 숨지게 한 '칠곡 계모사건'의 피고인 37살 임 모 씨에게 징역 15년을 선고했습니다. 또 임 씨가 의붓딸을 학대하도록 방조한 혐의로 ...

YTN | 2015.05.21. | 네이버뉴스
'칠곡 계모' 항소심서 징역 15년 선고받아 '살...' 매일경제 | 2015.05.21. | 네이버뉴스
법원 "칠곡 계모' 죄질이 무겁다" 항소심서 징... 매일경제 | 2015.05.21. | 네이버뉴스

연합뉴스 PiCK | 11시간 전 | 네이버뉴스
구미 사망 여아 친모 첫 공판...오늘 김천지원서
구미 3세 여아 사망 사건과 관련해 숨진 여아의 친모로 밝혀진 석모(48)씨에 대한 첫 공판이 22일 오전 여... 대구지법 김천지원 형사2단독 서청운 판사가 마성년자 약취 및 사체은...
"는 구미 사망 여아 친모... 국민일보 PiCK | 10시간 전 | 네이버뉴스
출산한 적 없어, 약취 전제... 이데일리 PiCK | 3시간 전 | 네이버뉴스
에 뿔난 엄마들 "친모 석씨..." 뉴시스 PiCK | 7시간 전 | 네이버뉴스
친모 첫 재판...진실 밝혀... 뉴시스 PiCK | 12시간 전 | 네이버뉴스

한스경제 | 7일 전
검찰, '정인이 사건' 양모 사형·양부 징역 7년 6개월 구형
연합뉴스 | 검찰이 16개월 된 입양아 정인아를 학대해 숨지게 한 양모에게 법정 최고형인 사형을... 앞서 장씨는 지난해 6월부터 10월까지 입양한 딸 정인양을 상습 폭행·학대하고, 그...
[1보] 검찰, 정인이 양모에 사형 구형..."반인륜적 범죄" 뉴스핌 | 7일 전
검찰 '정인이 사건' 양모 장 씨에 사형 구형 일요신문 | 7일 전
[종합] 정인이 양모에 사형 구형...양모 "나는 죽어 마땅" 눈물 뉴스핌 | 7일 전
[2보] 사형 구형되자 정인이 양모 눈물..."나는 죽어 마땅하다" 뉴스핌 | 7일 전

2020년 아동학대 신고 건수
“ 약 4만건 ”



01. 선정 배경 주제 소개

02. 전처리 및 EDA

03. 회귀 모델링

04. 다음주 예고

01. 주제 선정 배경

YTN | 2015.05.21. | 네이버뉴스
'칠곡 계모' 항소심 징역 15년
대구고등법원은 8살 난 의붓딸을 학대하고 폭행해 숨지게 한 '칠곡 계모사건'의 피고인 37살 임 모 씨에게 징역 15년을 선고했습니다. 또 임 씨가 의붓딸을 학대하도록 방조한 혐의로 ...

법원 "칠곡 계모' 죄질이 무겁다" 항소심서 징... 매일경제 | 2015.05.21. | 네이버뉴스

연합뉴스 PiCK | 11시간 전 | 네이버뉴스
구미 사망 여아 친모 첫 공판...오늘 김천지원서
구미 3세 여아 사망 사건과 관련해 숨진 여아의 친모로 밝혀진 석모(48)씨에 대한 첫 공판이 22일 오전 여... 대구지법 김천지원 형사2단독 서청운 판사가 마성년자 약취 및 사체은...
"는 구미 사망 여아 친모... 국민일보 PiCK | 10시간 전 | 네이버뉴스
출산한 적 없어, 약취 전제... 이데일리 PiCK | 3시간 전 | 네이버뉴스
에 뿔난 엄마들 "친모 석씨..." 뉴시스 PiCK | 7시간 전 | 네이버뉴스
친모 첫 재판...진실 밝혀... 뉴시스 PiCK | 12시간 전 | 네이버뉴스

한스경제 | 7일 전
검찰, '정인이 사건' 양모 사형·양부 징역 7년 6개월 구형
연합뉴스 | 검찰이 16개월 된 입양아 정인아를 학대해 숨지게 한 양모에게 법정 최고형인 사형을... 앞서 장씨는 지난해 6월부터 10월까지 입양한 딸 정인양을 상습 폭행·학대하고, 그...
[1보] 검찰, 정인이 양모에 사형 구형..."반인륜적 범죄" 뉴스핌 | 7일 전
검찰 '정인이 사건' 양모 장 씨에 사형 구형 일요신문 | 7일 전
[종합] 정인이 양모에 사형 구형...양모 "나는 죽어 마땅" 눈물 뉴스핌 | 7일 전
[2보] 사형 구형되자 정인이 양모 눈물..."나는 죽어 마땅하다" 뉴스핌 | 7일 전



2020년 아동학대 신고 건수
전년 대비 13.7% 증가

01. 주제 선정 배경

01. 선정 배경

주제 소개

02. 전처리 및 EDA

03. 회귀 모델링

04. 다음주 예고



크게 늘어나는 아동학대 건수에 비해
아동보호전문기관의 수는 현저히 적다!

01.
선정 배경

주제 소개

02.
전처리 및
EDA

03.
회귀 모델링

04.
다음주 예고

01. 주제 선정 배경

현재 예정된 정책 계획으로...

- 아동보호전문기관 10개 및 쉼터 29개 신규 설치
- 지자체 상담시설 설치 지원
- 21년 보건복지부 아동학대 대응 재원(국비) 437억원으로 작년 대비 140억원 증가(1.48배)

- 출처: 보건복지부

보건복지부에 따르면 2023년까지
아동보호전문기관 10개를 신규 설치할 예정



01. 주제 소개



서울시 아동보호전문기관 우선 입지 선정

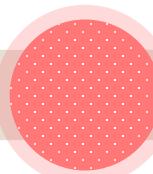
01.
선정 배경
주제 소개

02.
전처리 및
EDA

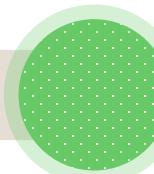
03.
회귀 모델링

04.
다음주 예고

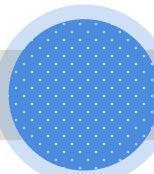
아동학대에 영향을
미치는 **변수 추출**



아동학대 발생 위험율이 높고
보호기관이 적은 행정동 선정



입지 조건을 만족하는
최종 입지 선정



아동보호전문기관의 효율적인 확충을 목표로 함!

01. 주제 소개



서울시 아동보호전문기관 우선 입지 선정

학대아동의 발견, 보호, 치료에 대한 신속한 처리 및
아동학대 예방을 위해 만들어진 전문 기관

- 아동학대 신고 접수 및 상담
- 현장출동 및 조사업무 수행
- 피해아동 및 가족에 대한 지원

아동학대와 관련된 전반적인 업무 수행

01.
선정 배경
주제 소개

02.
전처리 및
EDA

03.
회귀 모델링

04.
다음주 예고

01. 주제 소개



01.
선정 배경
주제 소개

02.
전처리 및
EDA

03.
회귀 모델링

04.
다음주 예고



서울시 아동보호전문기관 우선 입지 선정



2021년 기준

- 전국 69개
- 서울시 9개

운영 중

01. 주제 소개



서울시 아동보호전문기관 우선 입지 선정

01.
선정 배경
주제 소개

02.
전처리 및
EDA

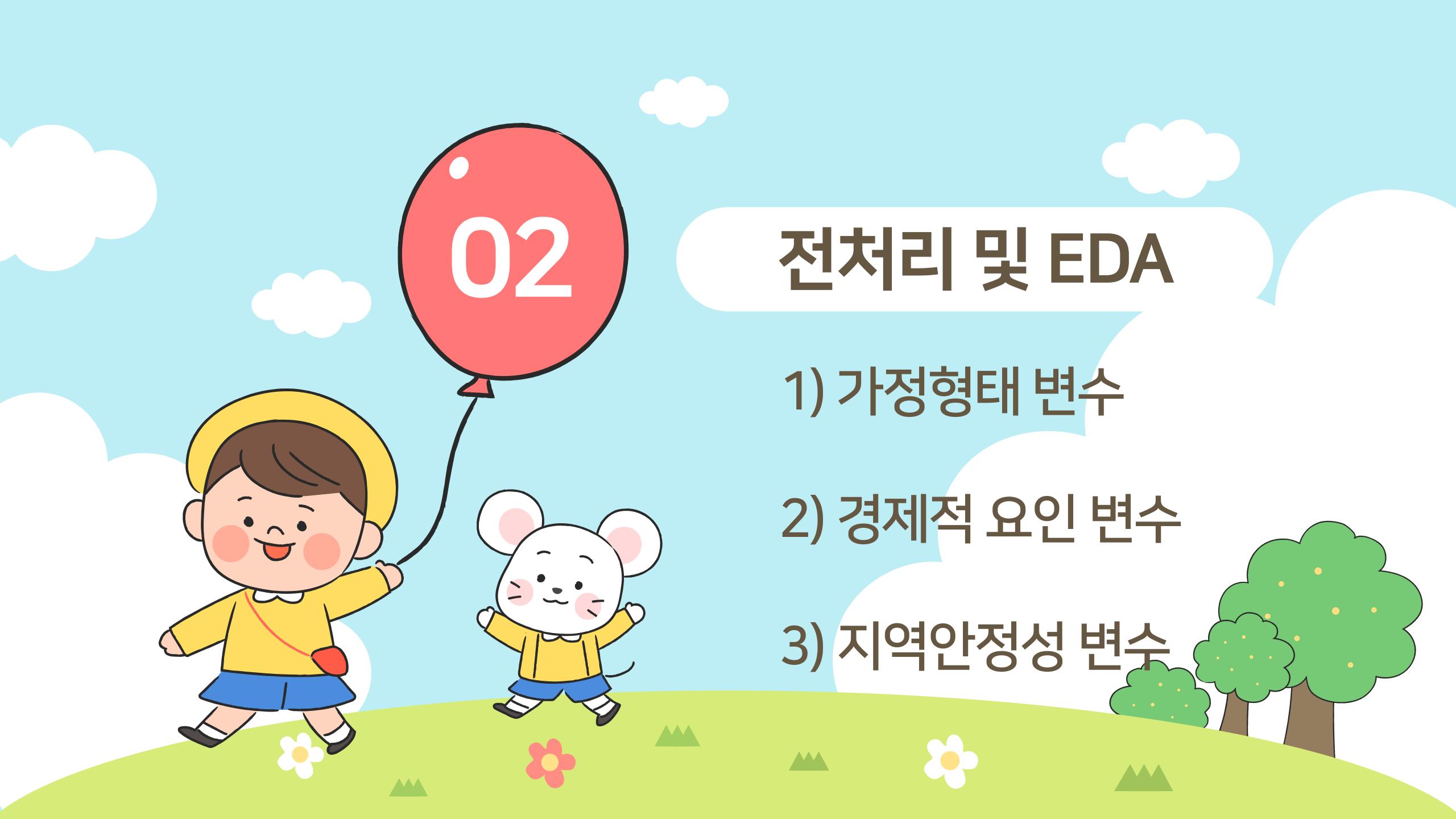
03.
회귀 모델링

04.
다음주 예고

잠깐!
아동의 기준이 뭔가요?

'아동'이란?

- UN아동권리협약에 의하면
18세 미만의 모든 사람을 의미!
- 우리나라 '아동복지법' 상에서도
만 18세 미만을 의미!



02

전처리 및 EDA

- 1) 가정형태 변수
- 2) 경제적 요인 변수
- 3) 지역안정성 변수

01.
주제 및
선정 배경

02.
가정형태 변수

경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 가정형태 변수

- 20.4분기 연령별 주민등록인구"에서 연령별 인구상세(동별).csv

동	성별	0세	1세	...	99세	100세+
종로구	계	469	585	...	44	276
	남	244	292	...	16	97
	여	225	293	...	28	179
사직동	계	26	37	...	2	21
	남	12	19	...	1	6
	여	14	18	...	1	15
삼청동	계	5	7	...	0	0

⋮

1354 obs x 108 variables



01.
주제 및
선정 배경

02.
가정형태 변수

경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 가정형태 변수

- 20.4분기 연령별 주민등록인구”에서 연령별 인구상세(동별).csv

동	성별	0세	1세	...	99세	100세+
종로구	계	469	585	...	44	276
	남	244	292	...	16	97
	여	225	293	...	28	179
사직동	계	26	37	...	2	21
	남	12	19	...	1	6
	여	14	18	...	1	15
삼청동	계	5	7	...	0	0

- 아동범위인 만 18세 인구까지 추출
- 성별과 관계없이 총합만 사용



01.
주제 및
선정 배경

02.
가정형태 변수

경제요인 변수
지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 가정형태 변수

- 20.4분기 연령별 주민등록인구”에서 연령별 인구상세(동별).csv

동	0세	...	18세
종로구	469	...	1,123
사직동	26	...	75
삼청동	5	...	17
부암동	26	...	96
평창동	70	...	175
무악동	38	...	111
종로구	469	...	1,123



동별아동수
16598
1139
259
1288
2416
1501

동	동별아동수
사직동	1139
삼청동	259
부암동	1288

구	구별아동수
종로구	16598
중구	12038
용산구	25975

만0세~18세 아동 수총합인 “동별아동수” 생성



01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수

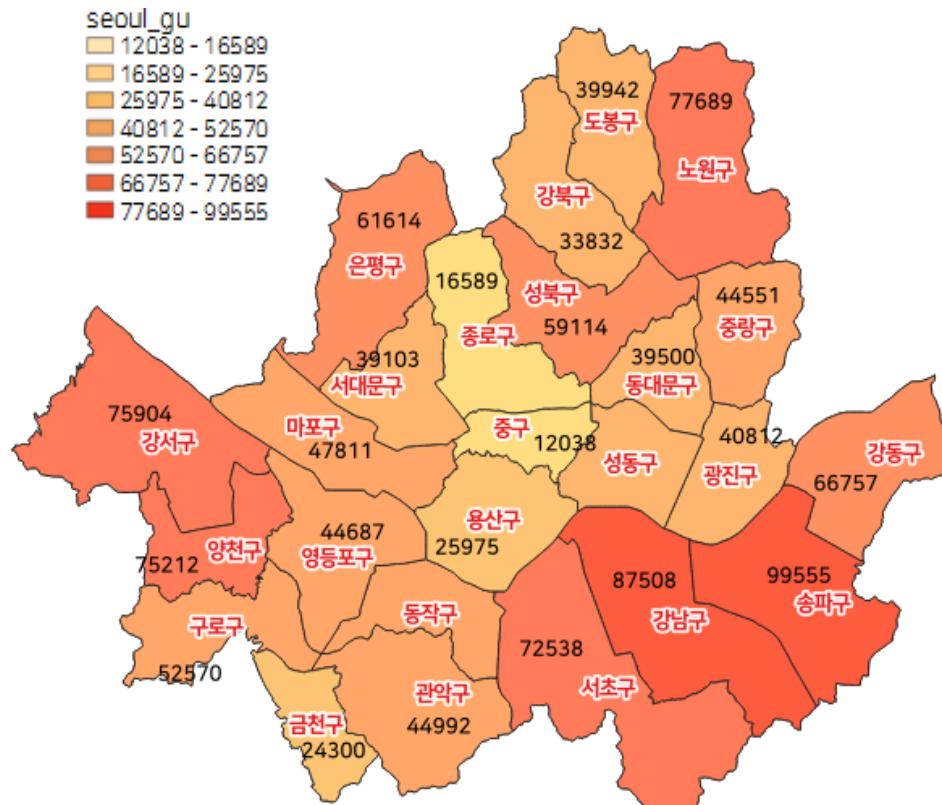
지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 가정형태 변수

- 서울특별시 구별 아동 인구수 시각화



✓ 아동 인구수가 **많은** 자치구
(강남구, 송파구, 노원구, 강서구, 서초구)

✓ 아동 인구수가 **적은** 자치구
(중구, 종로구, 금천구, 용산구, 강북구)



01.
주제 및
선정 배경

02.
가정형태 변수

경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

- 서울시 장애인 현황(연령별.동별).csv

기간	자치구	동	연령별	연령별	연령별	연령별	...	연령별
기간	자치구	동	합계	합계	합계	9세이하	...	80세이상
기간	자치구	동	계	남자	여자	계	...	여자
2019	합계	합계	394,843	228,821	166,022	4,771	...	35,259
2019	종로구	소계	6,068	3,515	2,553	55	...	653
2019	종로구	사직동	295	169	126	7	...	39
2019	종로구	삼청동	112	60	52	1	...	17
2019	종로구	부암동	316	179	137	6	...	47

455 obs x 33 variables



01.
주제 및
선정 배경

02.
가정형태 변수

경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 가정형태 변수

- 서울시 장애인 현황(연령별.동별).csv

기간	자치구	동	연령별	연령별	연령별	연령별	...	연령별
기간	자치구	동	합계	합계	합계	9세이하	...	80세이상
기간	자치구	동	계	남자	여자	계	...	여자
2019	합계	합계	394,843	228,821	166,022	4,771	...	35,259
2019	종로구	소계	6,068	3,515	2,553	55	...	653
2019	종로구	사직동	295	169	126	7	...	39
2019	종로구	삼청동	112	60	52	1	...	17
2019	종로구	부암동	316	179	137	6	...	47

자치구와 행정동 변수 사용

아동 인구인 '9세이하' & '19세이하' 사용



01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수
지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

- 서울시 장애인 현황(연령별.동별).csv

<구별 데이터>

자치구	구별장애인아동수
종로구	464
중구	254
용산구	536
성동구	732
광진구	732
동대문구	816
중랑구	1136
...	...

<동별 데이터>

동	동별장애인아동수
사직동	43
삼청동	4
부암동	24
평창동	30
무악동	36
교남동	30
가회동	9
...	...

424 obs. of 2 variables

포함되며 행정동이 424개...?

424개



01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 가정형태 변수

- 서울시 장애인 현황(연령별.동별).csv

서울시 전체 행정동 425개, 법정동 467개



01.
주제 및
선정 배경

02.
가정형태
변수

경제요인
변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 가정형태 변수

- 서울시 장애인 현황(연령별.동별).csv

해당지역을 클릭하시면 자세히 볼 수 있습니다.

전체	종로구	중구	용산구	성동구	광진구	동대문구	중랑구
성북구	강북구	도봉구	노원구	은평구	서대문구	마포구	양천구
강서구	구로구	금천구	영등포구	동작구	관악구	서초구	강남구
송파구	강동구						

17) 구로구(九老區) - 행정동(行政洞) 16, 법정동(法定洞) 10

行政洞	法定洞
신도림동(新道林洞)	신도림동(新道林洞)
구로제1동(九老第1洞)	구로동(九老洞)
구로제2동(九老第2洞)	구로동(九老洞)
구로제3동(九老第3洞)	구로동(九老洞)
구로제4동(九老第4洞)	구로동(九老洞)

하나하나 대조 끝에...



항동(航洞)

항동(航洞)

심은주 팀장님 Said :

“지난 주분 때 항동이 생길지 얼마 안돼서
데이터가 없다는 얘기를 들은 것 같기도.”

01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수
지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 가정형태 변수

- 서울시 장애인 현황(연령별.동별).csv

동	동별장애인수
사직동	43
삼청동	4
부암동	24
평창동	30
항동	0
교남동	30
가회동	9
...	...

항동을 추가해주고 값은 “0”으로 입력 후,
앞선 방법으로 전처리 진행

우리가 원하는
425개의 행정동 도출!



01.
주제 및
선정 배경

02.
가정형태
변수
경제요인
변수

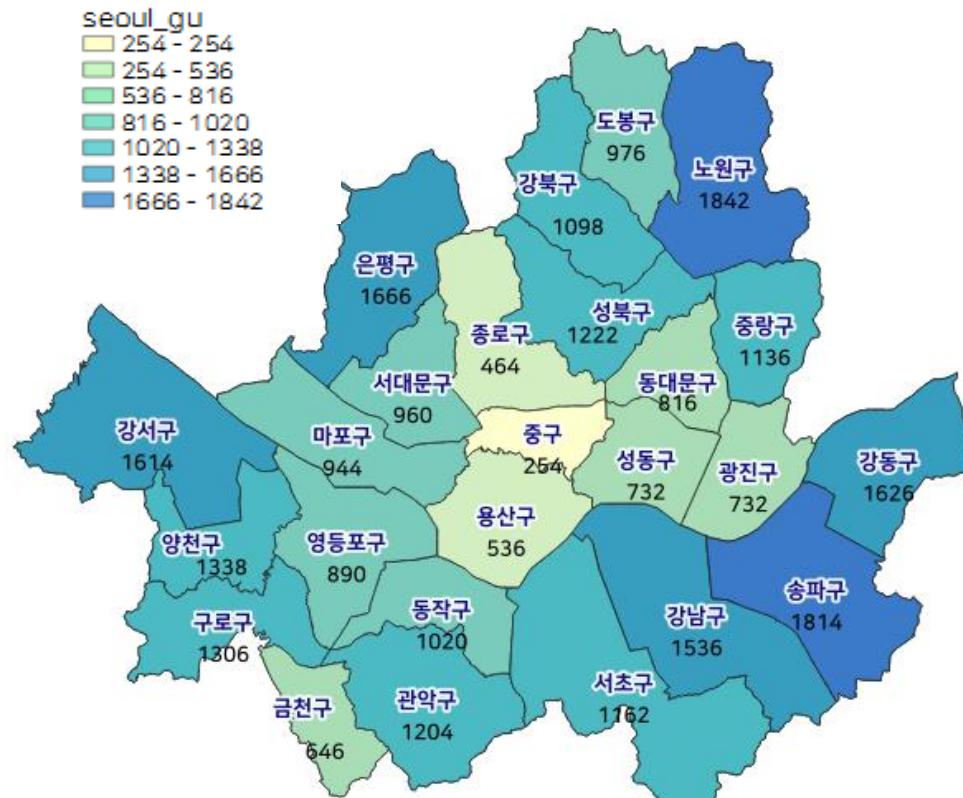
지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 가정형태 변수

- 서울특별시 구별 장애 아동수 시각화



장애 아동수가 많은 자치구
(노원구, 송파구, 은평구, 강동구, 강서구)

장애 아동수가 적은 자치구
(중구, 종로구, 용산구, 금천구, 광진구)



01.
주제 및
선정 배경

02.
가정형태 변수

경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

- 서울시 혼인이혼 (동별) 통계 (연간).xlsx

기간	자치구	동	혼인	이혼
2020	합계	합계	44,746	16,282
2020	종로구	소계	565	269
2020	종로구	사직동	40	19
2020	종로구	삼청동	22	5
2020	종로구	부암동	34	6
2020	종로구	평창동	58	24
2020	종로구	무악동	18	8

⋮

2260 obs x 5 variables



01.
주제 및
선정 배경

02.
가정형태
변수

경제요인
변수
지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 가정형태 변수

- 서울시 혼인이혼 (동별) 통계 (연간).xlsx

기간	자치구	동	혼인	이혼
2020	합계	합계	44,746	16,282
2020	종로구	소계	565	269
2020	종로구	사직동	40	19
2020	종로구	삼청동	22	5
2020	종로구	부암동	34	6
2020	종로구	평창동	58	24
2020	종로구	무악동	18	8

각 행정동 별 데이터만 사용!

이혼 데이터만 사용!



01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수
지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 가정형태 변수

- 서울시 혼인이혼 (동별) 통계 (연간).xlsx

자치구	동	이혼
합계	합계	16,282
종로구	소계	269
종로구	사직동	19
종로구	삼청동	5
⋮		

관악구	신사동	13
⋮		

강남구	신사동	9
⋮		

그런데
말입니다



서울에 두 개의 '신사동'이 존재!

(파주 토박이는 그런 거 몰라...)

Fix → 신사동_관

Fix → 신사동_강

01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수

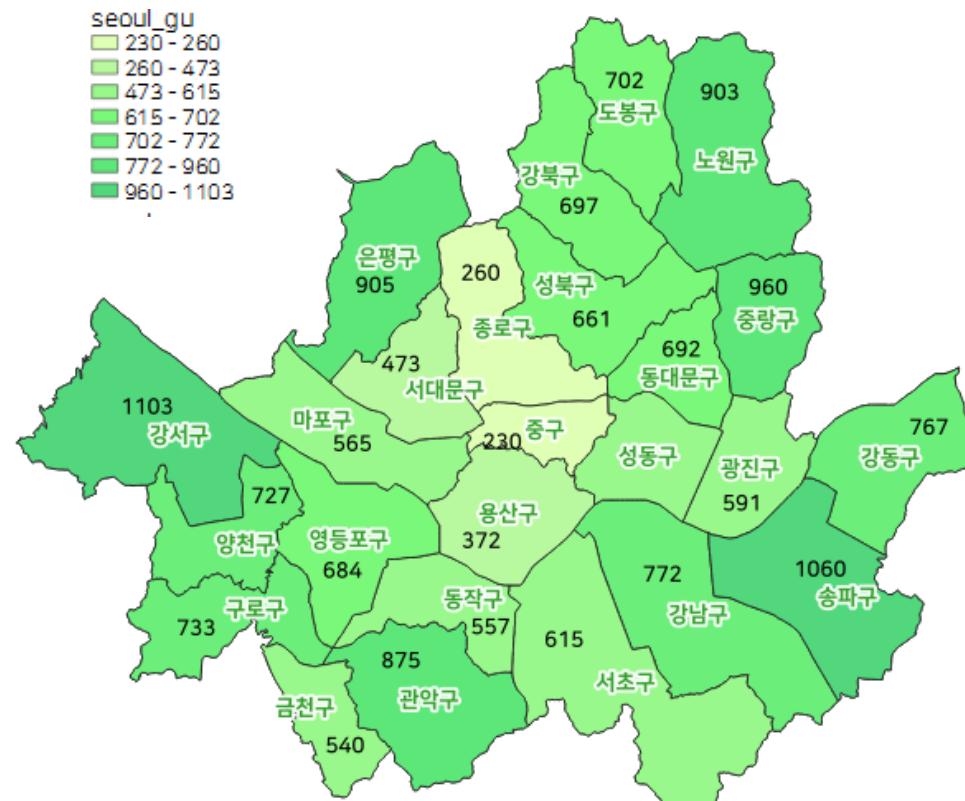
지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 가정형태 변수

- 서울특별시 구별 이혼 인구 시각화



이혼 인구가 **많은** 자치구
(강서구, 송파구, 중랑구, 은평구, 노원구)



이혼 인구가 **적은** 자치구
(중구, 종로구, 용산구, 금천구, 동작구)



01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 경제요인 변수

- 국토교통부 실거래가 공개시스템 (매매) 데이터 (2020년 1월-12월).csv

시군구	...	단지명	전용면적(m ²)	계약년월	계약일	거래금액(만원)	층	건축년도	도로명
서울특별시 강남구 개포동	...	개포2차현대 아파트(220)	77.75	202006	23	149,500	4	1988	언주로 103
서울특별시 강남구 개포동	...	개포6차우성 아파트1동 ~8동	54.98	202002	24	144,000	4	1987	언주로 3
서울특별시 강남구 개포동	...	개포6차우성 아파트1동 ~8동	79.97	202006	18	162,750	4	1987	언주로 3

(예시. 아파트 매매 데이터)

아파트 매매 실거래 데이터
14151 obs x 12 variables

단독/다가구 매매 실거래 데이터
12269 obs x 12 variables

연립/다세대 매매 실거래 데이터
62032 obs x 12 variables

오피스텔 매매 실거래 데이터
15210 obs x 13 variables

01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 경제요인 변수

- 국토교통부 실거래가 공개시스템 (매매) 데이터 (2020년 1월-12월).csv

시군구	...	단지명	전용면적(m ²)	계약년월	계약일	거래금액(만원)	층	건축년도	도로명
서울특별시 강남구 개포동	...	개포2차현대 아파트(220)	77.75	202006	23	149,500	4	1988	언주로 103
서울특별시 강남구 개포동	...	개포6차우성 아파트1동 ~8동	54.98	202002	24	144,000	4	1987	언주로 3
서울특별시 강남구 개포동	...	개포6차우성 아파트1동 ~8동	79.97	202006	18	162,750	4	1987	언주로 3

(예시. 아파트 매매 데이터)

주택 거래금액은 주택 크기에 영향을 받을 수 밖에 없음!

$$\text{단위가격} = \frac{\text{거래금액(만원)}}{\text{전용면적(m}^2\text{)}}$$

01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 경제요인 변수

- 국토교통부 실거래가 공개시스템 (매매) 데이터 (2020년 1월-12월).csv

시군구	...	단지명	전용면적(m ²)	계약년월	계약일	거래금액(만원)	층	건축년도	도로명
서울특별시 강남구 개포동	...	개포2차현대 아파트(220)	77.75	202006	23	149,500	4	1988	언주로 103
서울특별시 강남구 개포동	...	개포6차우성 아파트1동 ~8동	54.98	202002	24	144,000	4	1987	언주로 3
서울특별시 강남구 개포동	...	개포6차우성 아파트1동 ~8동	79.97	202006	18	162,750	4	1987	언주로 3

(예시. 아파트 매매 데이터)

정규표현식을 이용해 '자치구' 추출

더 정확한 도로명주소를 위해
'자치구'와 합쳐줌

강남구 + 언주로 103 = 강남구 언주로 103

01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 경제요인 변수

- 국토교통부 실거래가 공개시스템 (매매) 데이터 (2020년 1월-12월).csv

- 카카오 API를 통해 도로명 주소를 이용해 행정동 추출

도로명
주소

Hello,
WEB API

행정동

- 단독/다가구 주택의 경우, 도로명 주소가 'ROAD'로 뜨는 경우가 많아 위도&경도를 이용

도로명
주소

Hello,
WEB API

위도
&
경도

Hello,
WEB API

행정동

01.
주제 및
선정 배경

02.
가정형태
변수
경제요인
변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 경제요인 변수

- 국토교통부 실거래가 공개시스템 (매매) 데이터 (2020년 1월-12월).csv

<구별 데이터>

자치구	평균단위가격	거래 건수
강남구	2035.9725	3822
강동구	1192.4495	4230
강북구	704.9324	2177

<동별 데이터>

자치구	행정동	평균단위가격	거래 건수
강남구	개포1동	2659.1636	58
강남구	개포2동	2892.8175	184
강남구	개포4동	1668.9375	34

```

bow14[bowl4$도로명 == "강남구 남부순환로 2763", "행정동"] <- "도곡1동"
bow14[bowl4$도로명 == "강서구 방화동로 18-3", "행정동"] <- "공항동"
bow14[bowl4$도로명 == "강서구 광교대로 112", "행정동"] <- "화곡1동"
bow14[bowl4$도로명 == "광진구 능동로 37길 6-1", "행정동"] <- "중국1동"
bow14[bowl4$도로명 == "금천구 둑산로 353-1", "행정동"] <- "독산3동"
bow14[bowl4$도로명 == "금천구 시흥대로 75길 5", "행정동"] <- "시흥1동"
bow14[bowl4$도로명 == "마포구 만리재로 7길 2-3", "행정동"] <- "공덕동"
bow14[bowl4$도로명 == "서대문구 이화여대1길 21", "행정동"] <- "신촌동"
bow14[bowl4$도로명 == "서대문구 동교로 296-42", "행정동"] <- "연희동"
bow14[bowl4$도로명 == "성북구 정릉로 26길 48-3", "행정동"] <- "정릉2동"
bow14[bowl4$도로명 == "송파구 백제고분로 21길 7-1", "행정동"] <- "삼전동"
bow14[bowl4$도로명 == "은평구 진흥로 9길 3-3", "행정동"] <- "역촌동"
bow14[bowl4$도로명 == "중구 난계로 11길 18-4", "행정동"] <- "황학동"

```

재미있는 hand 크롤링을 통해 완성된
주택 유형별 구/동별 단위가격 데이터

그러나...

01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수

지역안전성
변수

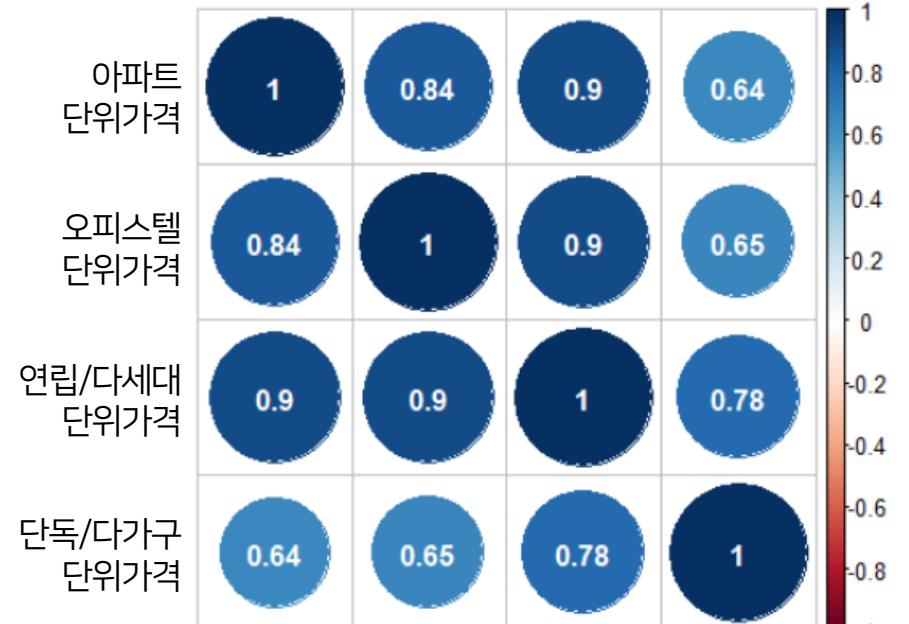
03.
회귀 모델링

04.
다음주 예고

02. 경제요인 변수

- 국토교통부 실거래가 공개시스템 (매매) 데이터 (2020년 1월-12월).csv

당연하게도(?) 주택 유형별 단위가격 간 **다중공선성이 매우 큼!**



부동산
단위가격

각 주택 유형별
거래 수

전체
매매 거래 수

각 주택 유형별
평균단위가격

의 합!

01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 경제요인 변수

- 국토교통부 실거래가 공개시스템 (매매) 데이터 (2020년 1월-12월).csv

<구별 데이터>

자치구	부동산 단위가격
강남구	1672.1418
강동구	942.5782
강북구	536.0634

<동별 데이터>

자치구	행정동	평균단위가격
강남구	개포1동	2659.164
강남구	개포2동	2741.674
강남구	개포4동	1010.519

자치구별/행정동별

부동산 가격 데이터 완성!

01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수

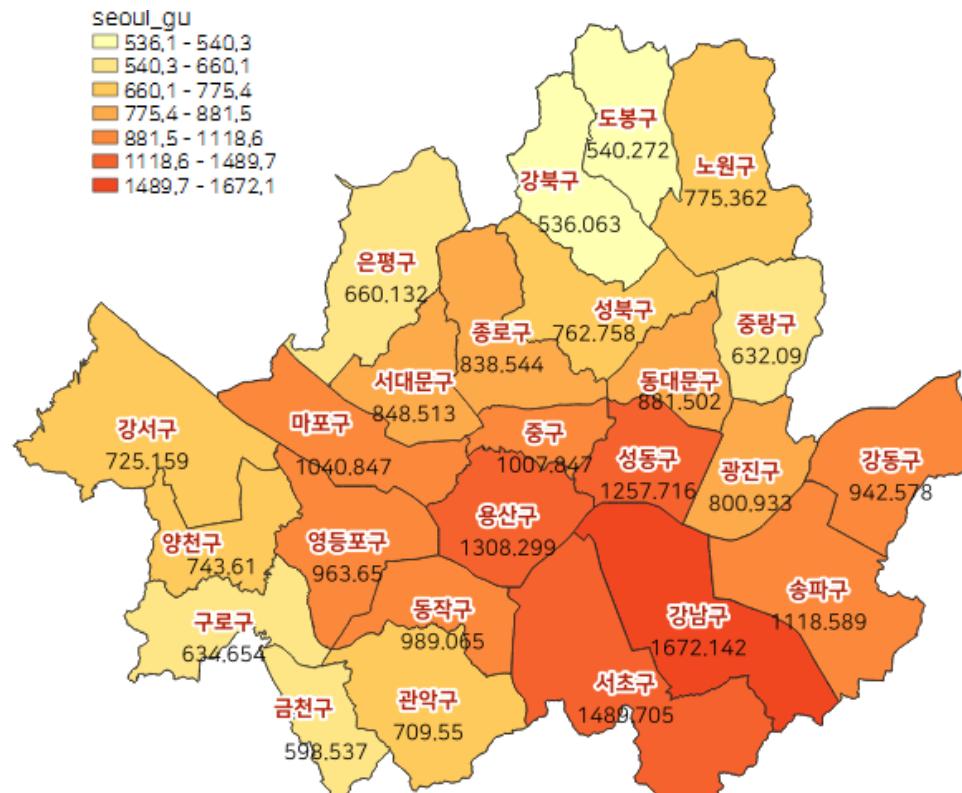
지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 경제요인 변수

- 서울특별시 구별 부동산 단위가격 시각화



부동산 가격이 **높은** 자치구
(강남구, 서초구, 용산구, 성동구, 송파구)

부동산 가격이 **낮은** 자치구
(강북구, 도봉구, 금천구, 중랑구, 구로구)



01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 경제요인 변수

- 서울시 국민기초생활보장 수급자 (동별) 통계.xlsx



기간	자치구	동	총수급자(가구)	총수급자(시설)	총수급자(인원)	...	시설수급자(인원)
2020	합계	합계	260,795	598	370,226	...	10,982
2020	본청	본청	-	5	708	...	708
2020	종로구	소계	3,851	13	5,232	...	609
2020	종로구	사직동	109	-	129	...	-
2020	종로구	삼청동	31	-	41	...	-
2020	종로구	부암동	114	-	150	...	-

476obs x 14 variables

01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 경제요인 변수

- 서울시 국민기초생활보장 수급자 (동별) 통계.xlsx



기간	자치구	동	총수급자(가구)	총수급자(시설)	총수급자(인원)	...	시설수급자(인원)
2020	합계	합계	260,795	598	370,226	...	10,982
2020	본청	본청	0	5	708	...	708
2020	종로구	소계	3,851	13	5,232	...	609
2020	종로구	사직동	109	-	129	...	-
2020	종로구	삼청동	31	-	41	...	-
2020	종로구	부암동	114	-	150	...	-



총 수급자 가구와 인원이 NA인 곳은 0으로 대체

01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 경제요인 변수

- 서울시 국민기초생활보장 수급자 (동별) 통계.xlsx

<구별 데이터>

자치구	총수급자(가구)_평균	총수급자(인원)_평균
강남구	426.9565	637.3913
강동구	572.3333	816.5556
강북구	1008.286	1417.571

⋮

<동별 데이터>

동	총수급자(가구)_평균	총수급자(인원)_평균
사직동	109	129
삼청동	31	41
부암동	114	150

⋮

BUT
총 수급자 가구와 인원의 경우, 해당 지역 **전체 인구의 영향**을
크게 받을 것이라고 생각해 수급 '**비율**'을 고려하기로 함!

01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 경제요인 변수

- 서울시 국민기초생활보장 수급자 (동별) 통계.xlsx



행정구	세대	내국인(계)
종로구	75,003	149,384
중구	63,686	125,240
용산구	113,093	230,040
성동구	135,870	293,556
광진구	167,427	346,682
동대문구	167,598	342,837

행정구	기초생활수급자 비율(가구)	기초생활수급자 비율(인원)
종로구	0.285248	0.194577
중구	0.415711	0.264442
용산구	0.323471	0.211344
성동구	0.288143	0.183913
광진구	0.316369	0.213885
동대문구	0.455534	0.29252

총수급자(가구)_평균/세대 * 100

총수급자(인원)_평균/내국인(계)* 100

01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 경제요인 변수

- 서울시 국민기초생활보장 수급자 (동별) 통계.xlsx

행정동	세대	내국인(계)
사직동	4,758	9,526
삼청동	1,342	2,640
부암동	4,392	10,022
평창동	7,522	18,391
무악동	3,192	8,546
교남동	4,602	10,477

행정동	기초생활수급자 비율(가구)	기초생활수급자 비율(인원)
사직동	2.290879	1.354189
삼청동	2.309985	1.55303
부암동	2.595628	1.496707
평창동	1.289551	0.74493
무악동	3.696742	1.813714
교남동	3.694046	2.090293

총수급자(가구)_평균/세대 * 100

총수급자(인원)_평균/내국인(계)* 100

01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 지역안전성 변수

- 경찰청_경찰관서 위치, 주소.csv

청	서	지구대파출소	X좌표	Y좌표
강원청	강릉경찰서	강동파출소	128.9783	37.72776
강원청	강릉경찰서	강릉경찰서	128.9068	37.76870
강원청	강릉경찰서	남부지구대	128.8971	37.74897
강원청	강릉경찰서	동부지구대	128.9263	37.77403
강원청	강릉경찰서	북부지구대	128.8752375	37.8352654
강원청	강릉경찰서	서부지구대	128.8755	37.7698333

2264 obs x 6 variables



01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 지역안전성 변수

- 경찰청_경찰관서 위치, 주소.csv

청	서	지구대파출소	X좌표	Y좌표
강원청	강릉경찰서	강동파출소	128.9783	37.72776
강원청	강릉경찰서	강릉경찰서	128.9068	37.76870
강원청	강릉경찰서	남부지구대	128.8971	37.74897

서울청 소속만 추출

274 obs x 4 variables

청	서	지구대파출소	주소
서울청	서울강남경찰서	논현1파출소	강남구 논현동 58-13
서울청	서울강남경찰서	논현3파출소	강남구 논현동 89-13
서울청	서울강남경찰서	삼성1파출소	강남구 삼성동 107-3

01.
주제 및
선정 배경

02.
가정형태
변수
경제요인
변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 지역안전성 변수

- 경찰청_경찰관서 위치, 주소.csv

청	서	지구대파출소	X좌표	Y좌표
강원청	강릉경찰서	강동파출소	128.9783	37.72776
강원청	강릉경찰서	강릉경찰서	128.9068	37.76870
강원청	강릉경찰서	남부지구대	128.8971	37.74897

위도 & 경도와 기관 주소를 이용해 자치구 및 행정동 추출

청	서	지구대파출소	주소	행정동	자치구
서울청	서울강남경찰서	논현1파출소	강남구 논현동 58-13	논현1동	강남구
서울청	서울강남경찰서	논현3파출소	강남구 논현동 89-13	논현2동	강남구
서울청	서울강남경찰서	삼성1파출소	강남구 삼성동 107-3	삼성1동	강남구

01.
주제 및
선정 배경

02.
가정형태
변수
경제요인
변수

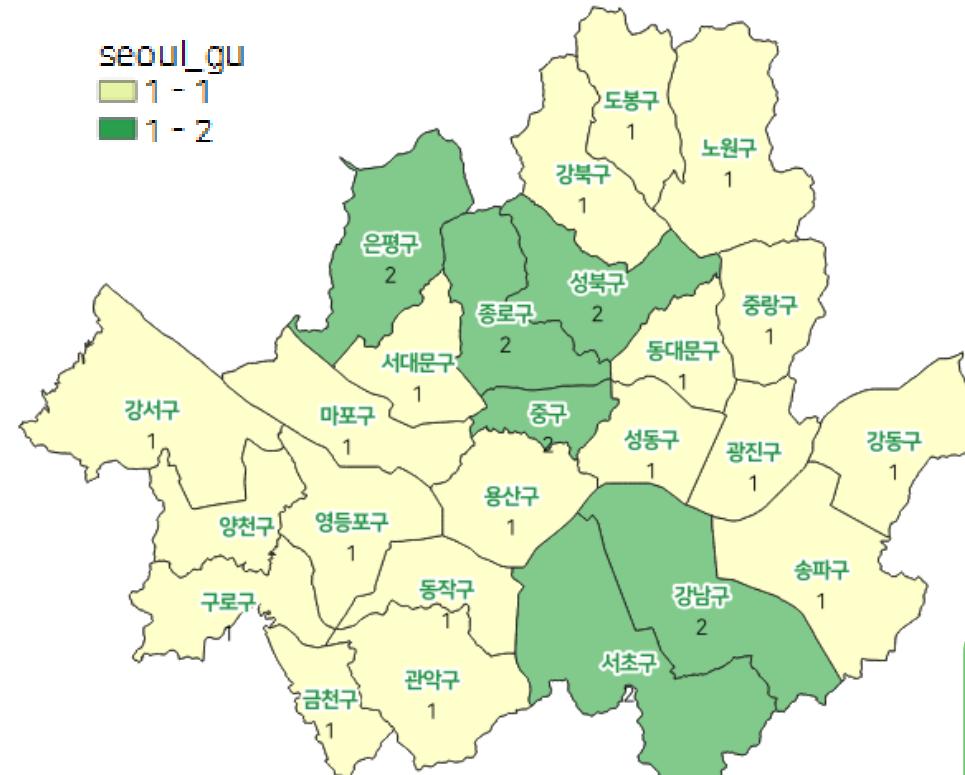
지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 지역안전성 변수

- 서울특별시 구별 경찰관서 수 시각화



자치구별 경찰관서 수의 차이는
보이지 않는다

대략적인 위치상
북쪽과 남쪽은 경찰서가
2개씩 있네!



01.
주제 및
선정 배경

02.
가정형태
변수
경제요인
변수

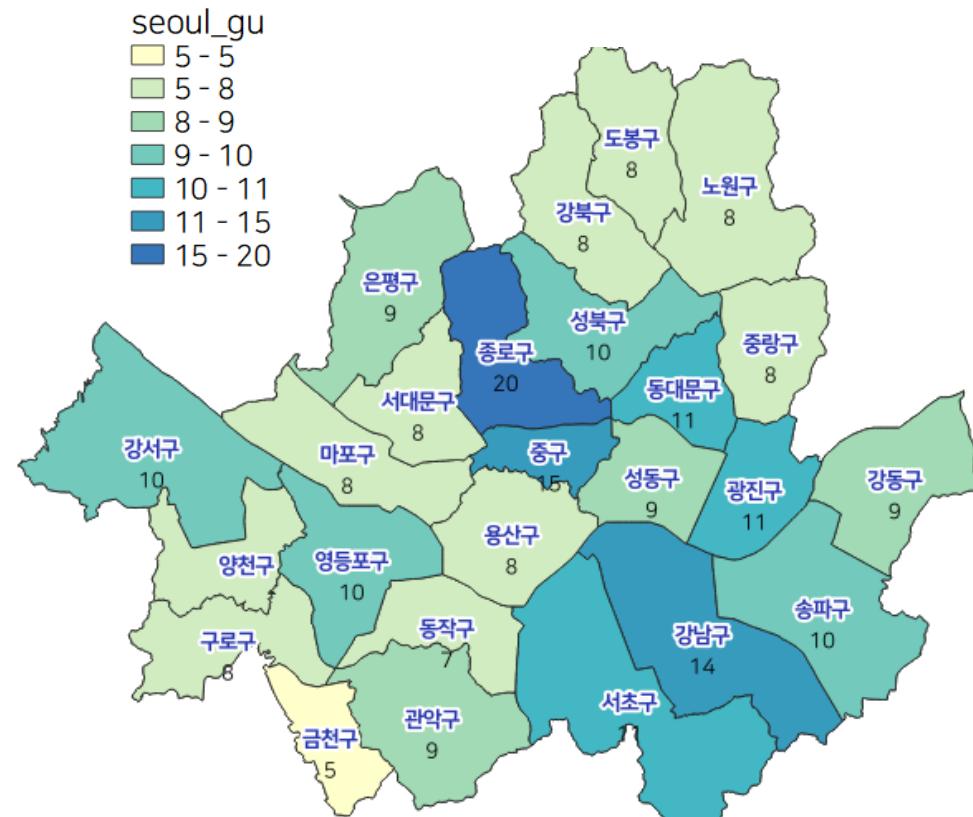
지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 지역안전성 변수

- 서울특별시 구별 지구대, 파출소 수 시각화



지구대, 파출소가 많은 자치구
(종로구, 중구, 강남구, 동대문구, 서초구)

지구대, 파출소가 적은 자치구
(금천구, 구로구, 양천구, 동작구, 용산구)



01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 지역안전성 변수

- 서울시 경찰관서별 가정폭력 데이터.csv

구분	검거건수	검거인원	구속	불구속
중부	86	66	1	97
종로	58	15	3	63
남대문	13	261	0	15
서대문	229	79	3	258
혜화	68	221	0	79
구분	검거건수	검거인원	구속	불구속

31 obs x 5 variables



01.
주제 및
선정 배경

02.
가정형태
변수
경제요인
변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 지역안전성 변수

- 서울시 경찰관서별 가정폭력 데이터.csv

구분	검거건수	검거인원	구속	불구속
중부	86	66	1	97
종로	58	15	3	63
남대문	13	261	0	15

각 경찰관서별 관할 자치구에
해당 검거건수를 배당하여
자치구별 검거건수 계산

자치구	검거건수
중구	86
종로구	58

01.
주제 및
선정 배경

02.
가정형태
변수
경제요인
변수

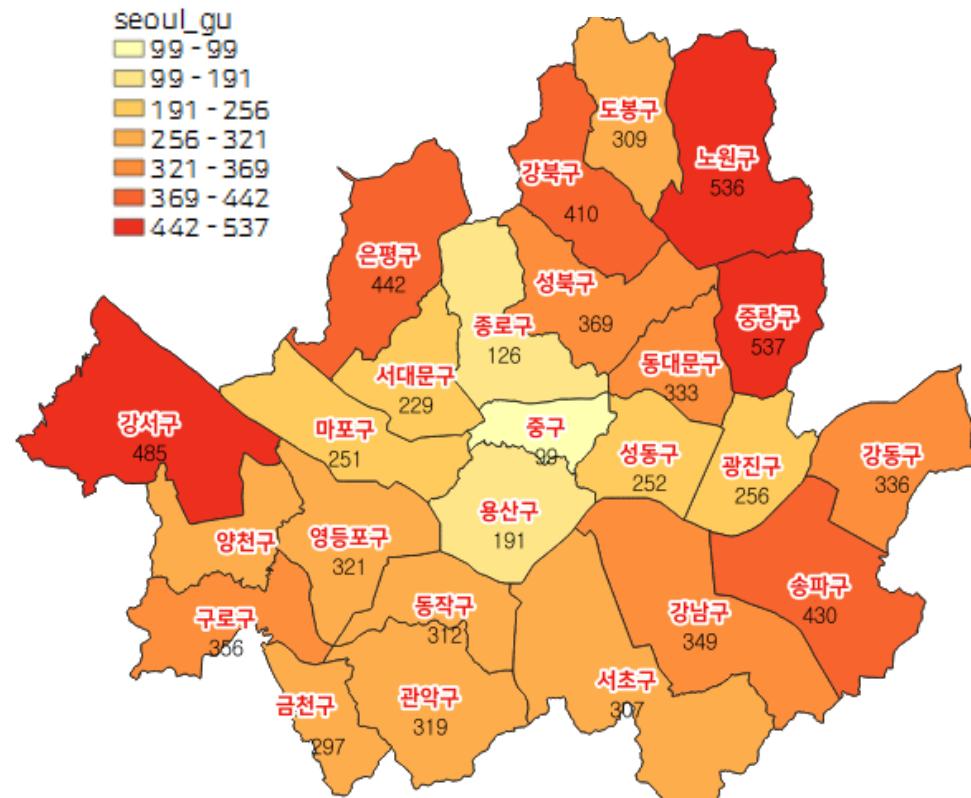
지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 지역안전성 변수

- 서울특별시 구별 가정폭력 검거건수 시각화



가정폭력이 **많은** 자치구
(중랑구, 노원구, 강서구, 은평구, 송파구)

가정폭력이 **적은** 자치구
(중구, 종로구, 서대문구, 마포구, 성동구)



01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 지역안전성 변수

- 서울시 5대 범죄 데이터.csv

기간	자치구	사건 발생 합계	사건 검거 합계	살인사건 발생	...
2019	종로구	3846	4117	4	...
2019	중구	4327	2804	2	...
2019	용산구	3313	2611	3	...
2019	성동구	2512	1838	6	...
2019	은평구	3567	1904	4	...

25 obs x 14 variables



01.
주제 및
선정 배경

02.
가정형태
변수
경제요인
변수

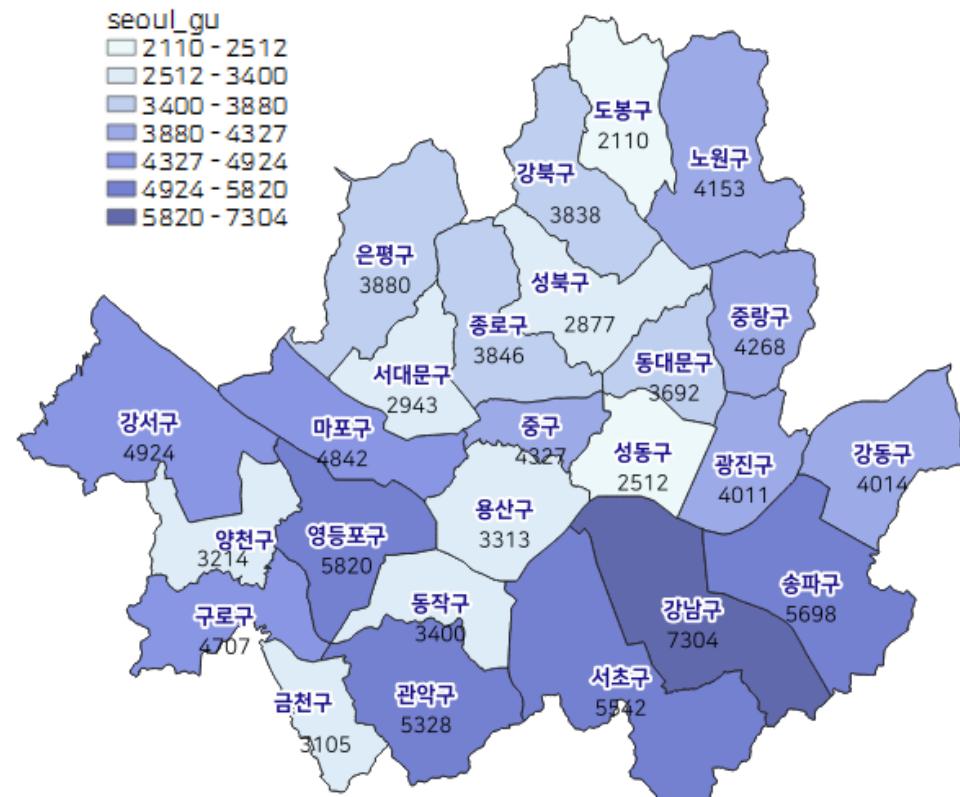
지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 지역안전성 변수

- 서울특별시 구별 5대 범죄건수 시각화



5대 범죄가 많은 자치구
(강남구, 영등포구, 송파구, 서초구, 관악구)

5대 범죄가 적은 자치구
(도봉구, 성동구, 성북구, 서대문구, 금천구)



02. 지역안전성 변수



- 서울특별시 구별 5대 범죄건수 시각화



5대 범죄가 많은 자치구 (강남구, 연등포구, 송파구, 서초구, 관악구) **행정동별 데이터를**

찾을 수 없었음

5대 범죄가 적은 자치구

사용 X

01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수
지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 지역안전성 변수

- 서울시 CCTV 데이터.csv

번호	관리기관명	소재지도로명주소	소재지지번주소	설치목적구분	카메라대수
1	서울특별시 중랑구청	서울특별시 중랑구 면 목로37길 58		생활방법	4
2	서울특별시 중랑구청	서울특별시 중랑구 용 마산로58길 35		생활방법	1
3	서울특별시 중랑구청	서울특별시 중랑구용 마산로62길 14		생활방법	3
4	서울특별시 광진구청		자양4동 199-4	생활방법	3
⋮					

37597 obs x 13 variables



01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수
지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 지역안전성 변수

- 서울시 CCTV 데이터.csv

번호	관리기관명	소재지도로명주소	소재지지번주소	설치목적구분	카메라대수
1	서울특별시 중랑구청	서울특별시 중랑구 면 목로37길 58		생활방법	4
2	서울특별시 중랑구청	서울특별시 중랑구 용 마산로58길 35		생활방법	1
3	서울특별시 중랑구청	서울특별시 중랑구용 마산로62길 14		생활방법	3
4	서울특별시 광진구청		자양4동 199-4	생활방법	3

설치목적 중 저희의 목적과 관련된
'생활방법', '어린이보호', '다목적', '기타' 만 필터링해 사용!



01.
주제 및
선정 배경

02.
가정형태
변수
경제요인
변수

지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 지역안전성 변수

- 서울시 CCTV 데이터.csv

소재지도로명주소	소재지지번주소
서울특별시 중랑구 면목로37길 58	
서울특별시 중랑구 용마산로58길 35	
	자양4동 199-4
	자양4동 199-4

시군구	행정동명
중랑구	면목7동
중랑구	면목7동
광진구	늑동
광진구	늑동

37597 obs
 $\times 15$ var

시군구, 행정동 정보 누락



카카오API를 사용해

- 소재지 도로명주소를 이용해 추출
- NA인 값에 대해 지번주소를 이용해 추출
- NA인 값에 대해 위도&경도를 이용해 추출

01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수
지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 지역안전성 변수

- 서울시 CCTV 데이터.csv

수많은 시도와...

눈물나는 hand 크롤링을 통해...

혹시...당신의 소중한 카카오 API KEY를 빌려줄 사람.....

혹시 카카오 API 키 있니

카카오 API 한번 더 지원받아요.....

한자로 쓰여진 성균관 표지석 앞 횡단보도 근처

최종 구별 / 동별 CCTV 데이터 완성

시군구	행정동명	카메라대수
중랑구	면목7동	4
중랑구	면목7동	1
광진구	늑동	3

01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수
지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 지역안전성 변수

- 서울시 CCTV 데이터.csv

BUT...

시군구	행정동명	카메라대수
중랑구	면목7동	4
중랑구	면목7동	1
광진구	늑동	3

눈물나는 hand 크롤링을 통해...

데이터의 행정동 개수 384개...

심은주

cctv가 없는 행정동이 존재할 수 있나...?

ㄴㄴㄴ 진수정

01.
주제 및
선정 배경

02.
가정형태 변수
경제요인 변수
지역안전성
변수

03.
회귀 모델링

04.
다음주 예고

02. 지역안전성 변수

- 서울시 CCTV 데이터.csv

BUT...

시군구	행정동명	카메라대수
중랑구	면목7동	4
중랑구	면목7동	1
광진구	늑동	3

384개

데이터의 행정동 개수 384개...

다음주에 다시 해보자^^!



03

회귀 모델링

- 1) 회귀 모델링
- 2) 위험도 지수
- 3) 대체 기관 수

01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

• 회귀모델링을 위한 최종 데이터셋

03. 회귀 모델링

25 obs X 8 var

자치구	장애아동수	아동수	이혼가정수	경찰서	지구대파출소	기초수급가구 비율	기초수급인구 비율	부동산 단위가격	아동학대 신고건수
종로구	464	16589	260	2	20	0.2852478	0.1945768	838.544	54
중구	254	12038	230	2	15	0.4157115	0.2644423	1007.8472	27
용산구	536	25975	372	1	8	0.3234714	0.2113443	1308.2993	96
성동구	732	35622	528	1	9	0.285	0.1839134	1257.7163	81
광진구	732	40812	591	1	11	0.3163692	0.2138848	800.9326	151

X

Y

아동학대 신고건수를 설명할 수 있는 변수들로 회귀모델링을 진행해,
베타 값들을 가중치로 두고 행정동별 '아동학대 위험 지수'를 산출

01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

03. 회귀 모델링

- 회귀모델링을 위한 최종 데이터셋

...	경찰서	지구대파출소	...
...	2	20	...
...	2	15	...
...	1	8	...
...	1	9	...
...	1	11	...

...	치안기관 수	...
...	22	...
...	17	...
...	9	...
...	10	...
...	12	...

경찰서 변수의 값이 1~2로 매우 작기 때문에
경찰서와 지구대·파출소의 합을 '치안기관 수' 변수로 사용

01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링

위험도 지수
대체기관 수

04.
다음주 예고

03. 회귀 모델링

- 상관계수, 다중공선성 확인

7개의 변수를 모두 사용한 회귀 모델에 대한 **VIF 값**

장애아동수	아동수	이혼가정수	치안기관 수	기초수급 가구비율	기초수급 인구비율	부동산 단위가격
12.773125	18.322310	7.545589	1.333304	102.33314	114.51064	2.842924

(**VIF > 10**)

다중공선성 존재 !

변수를 제거한
모델 비교

장애아동수, 기초수급인구비율

장애아동수, 기초수급가구비율

아동수, 기초수급인구비율

아동수, 기초수급가구비율

01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링

위험도 지수
대체기관 수

04.
다음주 예고

- 모델 비교

비교 모델	Model 1	Model 2	Model 3	Model 4
F-stat	0.004	0.005	0.004	0.005
R-square	0.465	0.4497	0.4627	0.4484

- 모든 회귀 모형이 통계적 유의성을 만족
- Model 1이 설명력이 가장 좋음

* Model 1 : (장애아동 수, 기초수급인구비율) 변수 제외
* Model 2 : (장애아동 수, 기초수급가구비율) 변수 제외
* Model 3 : (아동 수, 기초수급인구비율) 변수 제외
* Model 4 : (아동 수, 기초수급가구비율) 변수 제외

03. 회귀 모델링



01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

• 최종 모델

$$\text{아동학대 신고건수} = \text{아동수} + \text{이혼가정수} + \text{치안기관 수} + \text{기초수급 가구비율} + \text{부동산 단위가격}$$



F - Test: 모형의 유의성

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	45.2721355	98.0826277	0.462	0.650
아동수	0.0007411	0.0009843	0.753	0.461
이혼가정수	0.1079318	0.0939801	1.148	0.265
기초수급가구비율	118.9956765	101.7012886	1.170	0.256
부동산단위가격	-0.0709985	0.0503870	-1.409	0.175
경찰서파출소	0.0399523	3.3473908	0.012	0.991

Residual standard error: 46.97 on 19 degrees of freedom
Multiple R-squared: 0.5765, Adjusted R-squared: 0.465
F-statistic: 5.172 on 5 and 19 DF, p-value: 0.003668

P-value < 0.05

선택된 회귀 모형이
통계적으로 유의함을 확인!



01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

• 최종 모델

$$\text{아동학대 신고건수} = \text{아동수} + \text{이혼가정수} + \text{치안기관 수} + \text{기초수급 가구비율} + \text{부동산 단위가격}$$



F - Test: 모형의 유의성

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	45.2721355	98.0826277	0.462	0.650
아동수	0.0007411	0.0009843	0.753	0.461
이혼가정수	0.1079318	0.0939801	1.148	0.265
기초수급가구비율	118.9956765	101.7012886	1.170	0.256
부동산단위가격	-0.0709985	0.0503870	-1.409	0.175
경찰서파출소	0.0399523	3.3473908	0.012	0.991

Residual standard error: 46.97 on 19 degrees of freedom
Multiple R-squared: 0.5765, Adjusted R-squared: 0.465
F-statistic: 5.172 on 5 and 19 DF, p-value: 0.003668

P-value < 0.05

선택된 회귀 모형이
통계적으로 유의함을 확인!!



03. 회귀 모델링



01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

- 최종 모델

$$\text{아동학대 신고건수} = \text{아동수} + \text{이혼가정수} + \text{치안기관 수} + \text{기초수급 가구비율} + \text{부동산 단위가격}$$



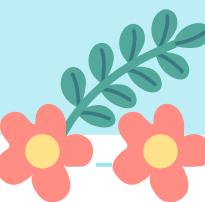
회귀 가정 확인

- Gvlma 패키지

```
call:  
gvlma(x = lr_model1)
```

	value	p-value	Decision
Global stat	5.1800	0.2693	Assumptions acceptable.
Skewness	1.0061	0.3158	Assumptions acceptable.
Kurtosis	0.1564	0.6925	Assumptions acceptable.
Link Function	1.4678	0.2257	Assumptions acceptable.
Heteroscedasticity	2.5497	0.1103	Assumptions acceptable.

모든 회귀 가정들이
충족됨을 확인!



03. 회귀 모델링



01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

- 최종 모델

$$\text{아동학대 신고건수} = \text{아동수} + \text{이혼가정수} + \text{치안기관 수} + \text{기초수급 가구비율} + \text{부동산 단위가격}$$



회귀 가정 확인

- Gvlma 패키지

```
call:  
gvlma(x = lr_model1)
```

	value	p-value	Decision
Global stat	5.1800	0.2693	Assumptions acceptable.
Skewness	1.0061	0.3158	Assumptions acceptable.
Kurtosis	0.1564	0.6925	Assumptions acceptable.
Link Function	1.4678	0.2257	Assumptions acceptable.
Heteroscedasticity	2.5497	0.1103	Assumptions acceptable.

모든 회귀 가정들이
충족됨을 확인!



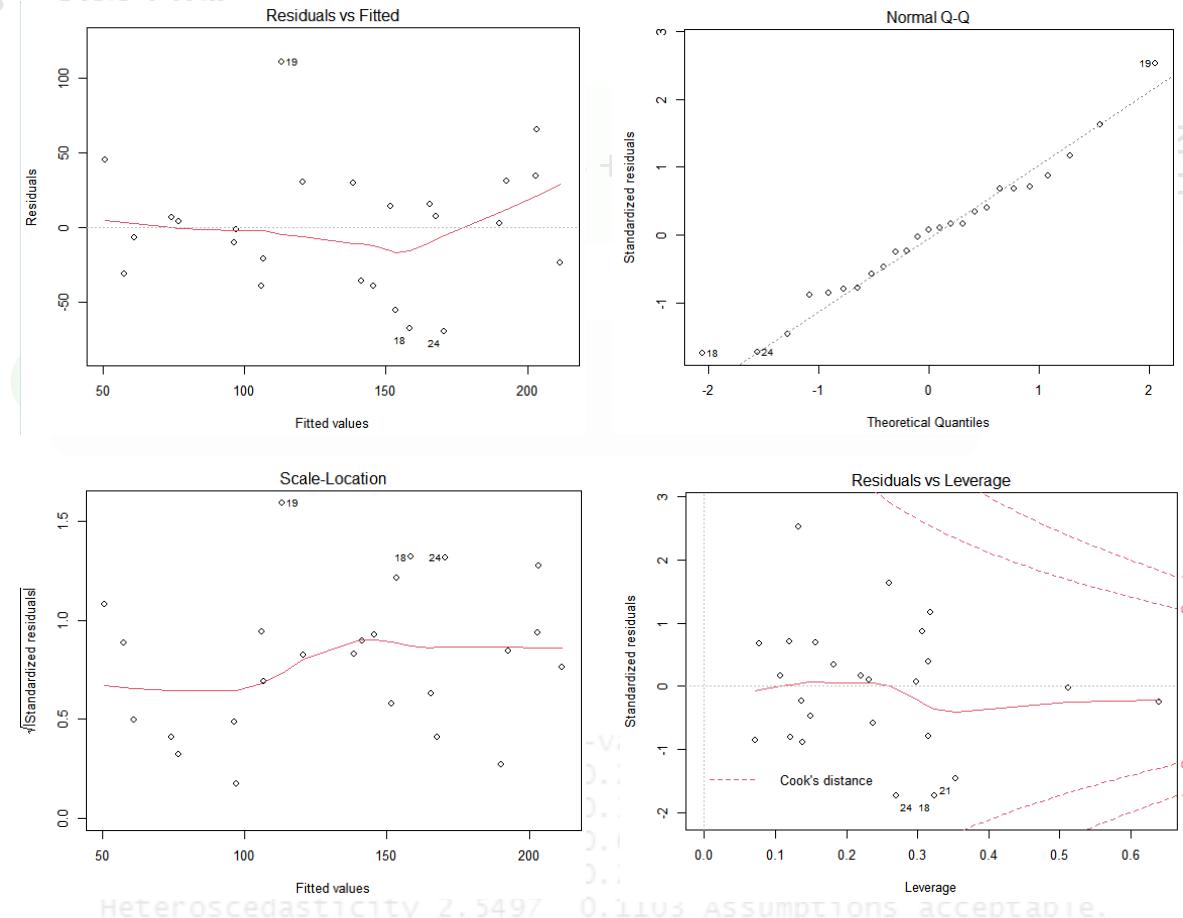
01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

03. 회귀 모델링



초수급
구비율
+
부동산
단위가격

잔차 플랏으로도

회귀 가정들을 만족하는 것을

모든 ~~회귀~~ 가정이
확인 가능!

충족됨을 확인!

01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

• 최종 모델

$$\text{아동학대 신고건수} = \text{아동수} + \text{이혼가정수} + \text{치안기관 수} + \frac{\text{기초수급 가구비율}}{} + \frac{\text{부동산 단위가격}}{}$$



다중공선성

아동 수	이혼가정 수	치안기관 수	기초수급가구비율	부동산단위가격
5.144047	4.652090	1.329626	2.280880	2.331491

모든 변수의 VIF 값이 10보다 작음

→ 다중공선성 없음!

03. 회귀 모델링



01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

• 최종 모델

$$\text{아동학대 신고건수} = \text{아동수} + \text{이혼가정수} + \text{치안기관 수} + \frac{\text{기초수급 가구비율}}{\text{부동산 단위가격}}$$



다중공선성

아동 수	이혼가정 수	치안기관 수	기초수급가구비율	부동산단위가격
5.144047	4.652090	1.329626	2.280880	2.331491

모든 변수의 VIF 값이 10보다 작음



다중공선성 없음!



01.
주제 및
선정 배경

02.
전처리 및
EDA

03.

회귀 모델링
위험도 지수

대체기관 수

04.
다음주 예고

03. 위험도 지수

- 회귀계수를 사용한 위험도 지수 계산

아동학대 신고건수 =

$$\beta_0 + \beta_1 \times \text{아동 수} + \beta_2 \times \text{이혼 가정 수} + \beta_3 \times \text{치안기관 수} \\ + \beta_4 \times \text{기초수급가구 비율} + \beta_5 \times \text{부동산 단위가격}$$

→ 회귀계수를 가중치로 두고 행정동별 위험도 지수 산출

위험도 지수 =

$$\beta_1 \times \text{아동 수} + \beta_2 \times \text{이혼 가정 수} + \beta_3 \times \text{치안기관 수} \\ + \beta_4 \times \text{기초수급가구 비율} + \beta_5 \times \text{부동산 단위가격}$$



01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수

대체기관 수
04.
다음주 예고

03. 위험도 지수

- 회귀계수를 사용한 위험도 지수 계산

BUT

자치구	아동수	이혼가정수	치안기관 수	기초수급가구비율	부동산 단위가격
종로구	16589	260	22	0.2852478	838.544

변수별 단위의 차이가 너무 큼!

$$\text{표준화계수} = \text{비표준화 계수} \times \frac{S_x}{S_y}$$

: 독립변수의 표준편차
: 종속변수의 표준편차

01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

03. 위험도 지수

- 회귀계수를 사용한 위험도 지수 계산

행정동	아동수	이혼가정수	치안기관 수	기초수급가구비율	부동산 단위가격
사직동	1139	13	4	2.2908	1016.3605
부암동	259	9	2	2.5956	668.9841
평창동	1288	16	2	1.2895	630.3685

동별 데이터를 이용해 행정동별 위험도 지수 산출!

위험도 지수 =

β_i^* : 표준화계수

$$\beta_1^* \times \text{아동 수} + \beta_2^* \times \text{이혼 가정 수} + \beta_3^* \times \text{치안기관 수} \\ + \beta_4^* \times \text{기초수급가구 비율} + \beta_5^* \times \text{부동산 단위가격}$$

01.
주제 및
선정 배경

02.
전처리 및
EDA

03.

회귀 모델링
위험도 지수

대체기관 수

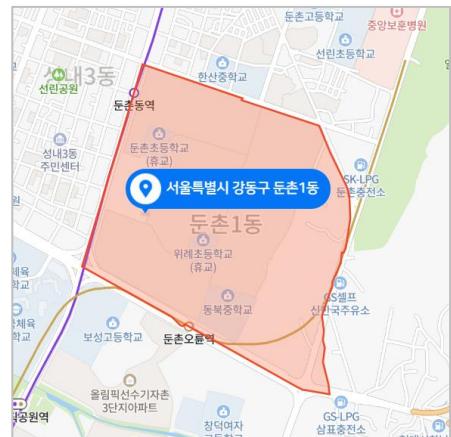
04.
다음주 예고

03. 위험도 지수

- 회귀계수를 사용한 위험도 지수 계산

그런데... ↗?

행정동	아동수	이혼가정수	치안기관 수	기초수급가구비율	부동산 단위가격
삼청동	259	9	2	2.3099	0
가회동	474	5	0	2.6480	0
둔촌1동	9	6	0	0	0



단위기간동안 매매거래가 없었던 3개의 행정동

위험도 지수 산출에 미치는 영향이 크다고 판단해
'네이버 부동산'에 등록된 정보로 계산해 입력!

→ 다만 '둔촌1동'의 경우 주거지가 거의 없어, 매매 매물 자체가 없음!



01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수

대체기관 수

04.
다음주 예고

03. 위험도 지수

- 회귀계수를 사용한 위험도 지수 계산

위험도 지수 데이터 완성!

행정동	아동수	이혼가정수	치안기관수	기초수급가구비율	부동산단위가격	위험도지수
사직동	1139	13	4	2.29087852	1016.360556	-49.348256
삼청동	259	9	2	2.309985097	43760.57892	-7348.0277
부암동	1288	16	2	2.595628415	668.9841244	220.901380

⋮

425 obs x 7 variables

01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

03. 대체기관 수 변수

- 아동청소년 보호기관 정보.csv

기관명	시도	시군구	읍면동	행정동명	기관주소	...
도담 상대원동 지역아동센터	서울특별시	노원구	상계동	상계5동	서울특별시 노원구 한글비석로 460	...
11월의혁명 보습학원	서울특별시	양천구	신정동	신정4동	서울특별시 양천구 오목로50길 33	...
가락중학교	부산광역시	강서구	죽림동	가락동	부산광역시 강서구 가락대로 1405	...
강일늘사랑 유치원	서울특별시	강동구	강일동		서울특별시 강동구 아리수로 98길 62	...

:

144229 obs x 24 variables



01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

03. 대체기관 수 변수

- 아동청소년 보호기관 정보.csv

기관명	시도	시군구	읍면동	행정동명	기관주소	...
도담 상대원동 지역아동센터	서울특별시	노원구	상계동	상계5동	서울특별시 노원구 한글비석로 460	...
11월의혁명 보습학원	서울특별시	양천구	신정동	신정4동	서울특별시 양천구 오목로50길 33	...
가락중학교	부산광역시	강서구	죽림동	가락동	부산광역시 강서구 가락대로 1405	...
강일늘사랑 유치원	서울특별시	강동구	강일동	NA	서울특별시 강동구 아리수로 98길 62	...

일부 행정동 변수에 NA 값이 포함되어 있어



카카오 API를 이용해
기관주소로 행정동 추출

NA 처리

성내1동

서울특별시 강동구 아리수로 98길 62

01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

03. 대체기관 수 변수

- 아동청소년 보호기관 정보.csv

기관명	시도	시군구	읍면동	행정동명	기관주소	...
도담 상대원동 지역아동센터	서울특별시	노원구	상계동	상계5동	서울특별시 노원구 한글비석로 460	...
11월의혁명 보습학원	서울특별시	양천구	신정동	신정4동	서울특별시 양천구 오목로50길 33	...
가락중학교	부산광역시	강서구	죽림동	가락동	부산광역시 강서구 가락대로 1405	...
강일늘사랑 유치원	서울특별시	강동구	강일동		서울특별시 강동구 아리수로 98길 62	...

Filtering
& Selecting

기관명	시군구	행정동명
도담 상대원동 지역아동센터	노원구	상계5동
강일늘사랑 유치원	강동구	성내 1동

서울특별시가 아닌 행,
보호기관으로 적절하지 않은 행(학원, 독서실 등) 존재

7993 obs X 3 variables

01.
주제 및
선정 배경

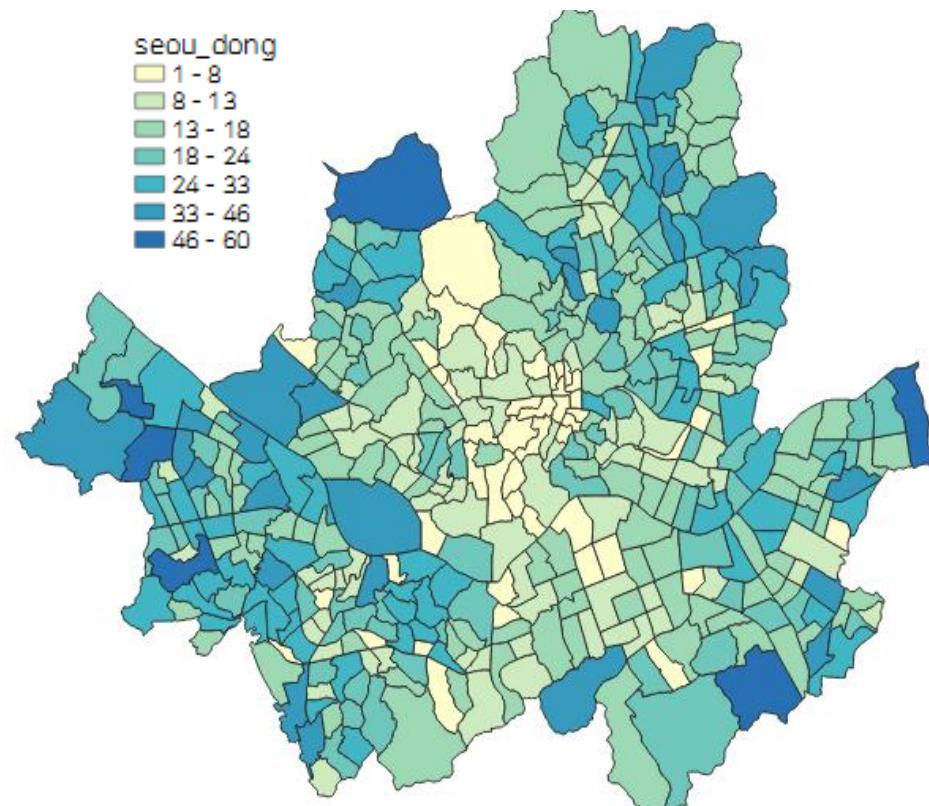
02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

03. 대체기관 수 변수

- 서울특별시 동별 아동청소년 보호기관 수 시각화



✓ 아동청소년보호기관수가 많은 행정동
(발산1동, 신정3동, 방화1동, 진관동, 강일동)

✓ 아동청소년보호기관수가 적은 행정동
(개포1동, 둔촌1동, 광희동, 장충동, 을지로동)



01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

03. 대체기관 수 변수

- 서울시 가정폭력상담소 정보.xlsx

사업장명	소재지전체주소	도로명 전체주소	인허가일자	영업상태명	폐업일자	...
나우리가족 상담센터	서울특별시 마포구 성산동 592번지		20070320	운영중		...
용산가정폭력 관련상담소	서울특별시 용산구 효창동 5-575번지		20011016	운영중		...
사)남성의전화 부설 서울가정폭력상담센터	서울특별시 양천구 목동 917-9번지		20050725	운영중		...

33 obs x 19 variables

01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

03. 대체기관 수 변수

- 서울시 가정폭력상담소 정보.xlsx

사업장명	소재지전체주소
나우리가족 상담센터	서울특별시 마포구 성산동 592번지
용산가정폭력 관련상담소	서울특별시 용산구 효창동 5-575번지
사)남성의전화 부설 서울가정폭력상담센터	서울특별시 양천구 목동 917-9번지

시군구	행정동
마포구	성산2동
용산구	효창동
양천구	목1동



소재지 전체주소로부터 시군구 정보 추출

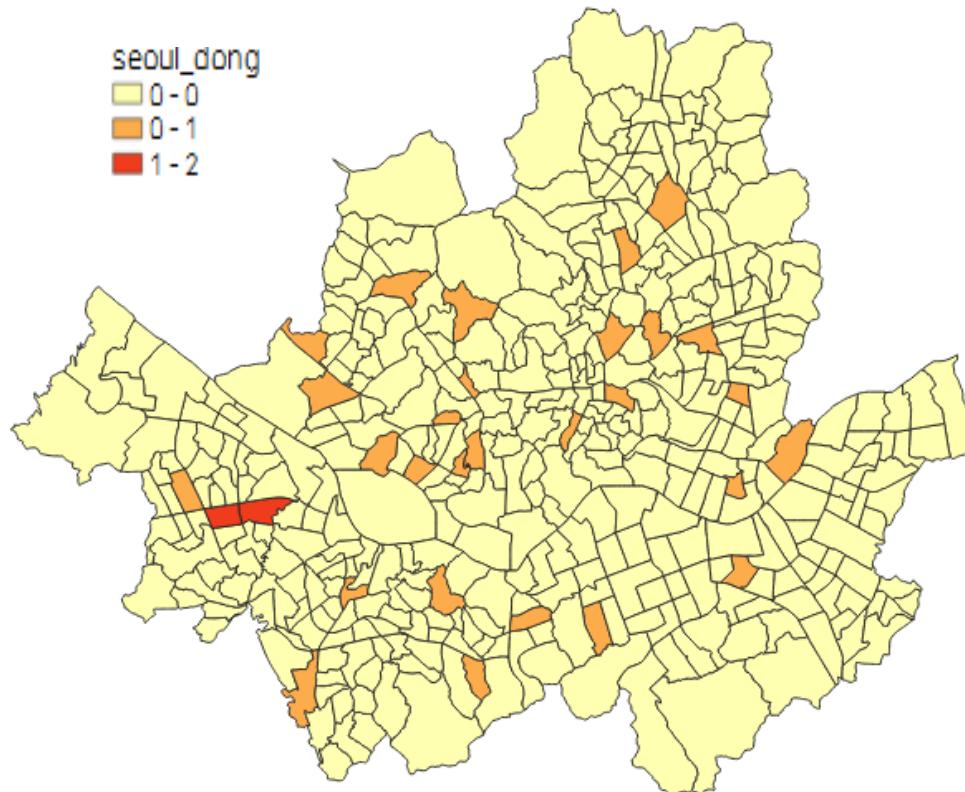
카카오 API를 활용하여 행정동명 맵핑

32 obs x 5 variables

03. 대체기관 수 변수



- 서울특별시 동별 가정폭력상담소 수 시각화



01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

가정폭력상담소의 수가 적어
행정동별로 유의미한 차이를
보이지는 않음

크게 밀집되어 있는 지역은 없지만,
강북과 강남 방향 외곽에는 많이 없음

01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

03. 대체기관 수 변수

- 서울특별시 아동복지시설 목록.csv

시설명	시설코드	시설유형	시설종류	자치구	시설장명	시군구코드
구세군서울 후생원	B0001	아동양육시설	아동복지시설	자치구	김호규	1141000000
송파희망세상지역아동센터	K0218	지역아동센터	아동복지시설	자치구	이정훈	1171000000
송파희망세상지역아동센터	K1580	지역아동센터	아동복지시설	자치구	이정훈	1171000000
한국아동복지시설연합회	B0704	아동양육시설	아동복지시설	자치구	성준경	1100000000
조대봉님	P0002	아동양육시설	아동복지시설	자치구	조대봉	1165000000

⋮

650 obs x 9 variables



01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

03. 대체기관 수 변수

- 서울특별시 아동복지시설 목록.csv

시설명	시설코드	시설유형	시설종류	자치구	시설장명	시군구코드
구세군서울 후생원	B0001	아동양육시설	아동복지시설	자치구	김호규	1141000000
송파희망세상지역아동센터	K0218	지역아동센터	아동복지시설	자치구	이정훈	1171000000
송파희망세상지역아동센터	K1580	지역아동센터	아동복지시설	자치구	이정훈	1171000000
한국아동복지시설연합회	B0704	아동양육시설	아동복지시설	자치구	성준경	1100000000
조대봉님	P0002	아동양육시설	아동복지시설	자치구	조대봉	1165000000

시설이 중복된 행 /
아동복지시설이 아닌 행 /
주소지가 서울이 아닌 행 존재

Filtering & Selecting

시설명	시군구명	시설주소
구세군서울 후생원	서대문구	서울특별시 서대문구 독립문로8길 41
송파희망세상지역아동센터	송파구	서울특별시 송파구 오금로64길 4 2층 (거여동)
한국아동복지시설연합회	서울특별시	서울특별시 용산구 갈월동 용산빌딩

01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

03. 대체기관 수 변수

- 서울특별시 아동복지시설 목록.csv

시설명	시군구명	시설주소
구세군서울 후생원	서대문구	서울특별시 서대문구 독립문로8길 41
송파희망세상 지역아동센터	송파구	서울특별시 송파구 오금로64길 4 2층 (거여동)
한국아동복지 시설연합회	서울특별시	서울특별시 용산구 갈월동 용산빌딩

- 잘못된 시군구 존재
- 행정동명 정보 누락

650 obs X 9 variables



주소 데이터를 활용하여 잘못된 시군구 변경

카카오 API를 활용하여 행정동명 변수 추가

시설명	시군구명	시설주소	행정동명
구세군서울 후생원	서대문구	서울특별시 서대문구 독립문로8길 41	천연동
송파희망세상 지역아동센터	송파구	서울특별시 송파구 오금로64길 4 2층 (거여동)	거여2동
한국아동복지 시설연합회	용산구	서울특별시 용산구 갈월동 용산빌딩	공덕동

01.
주제 및
선정 배경

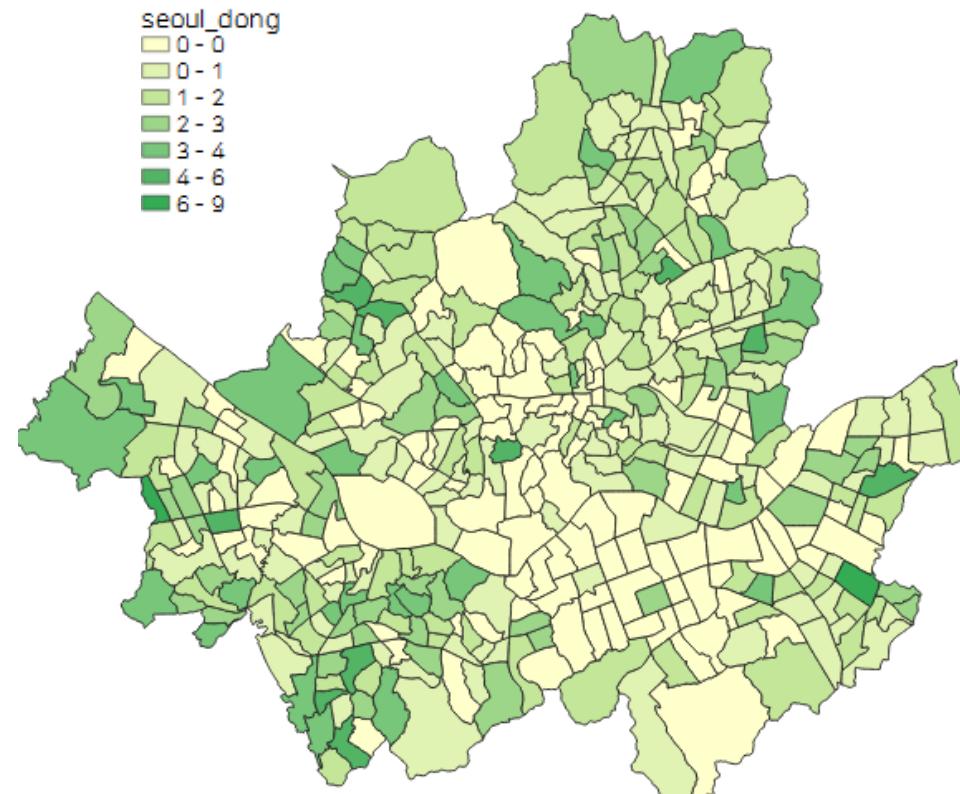
02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

03. 대체기관 수 변수

- 서울특별시 동별 아동복지시설 수 시각화



✓ 아동복지시설이 **많은** 행정동
(신월3동, 오금동, 오류2동, 미성동, 길동)

✓ 아동복지시설이 **적은** 행정동
(개포1동, 능동, 무악동, 명동, 보문동 등)



01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링
위험도 지수
대체기관 수

04.
다음주 예고

03. 대체기관 수 변수

- 클러스터링을 위한 최종 데이터셋

행정동	위험도지수	대체기관 수
사직동	-49.348256	1024
삼청동	-7348.0277	114
부암동	220.901380	436
평창동	516.5103416	792
무악동	385.639566	593
교남동	144.3281057	797
가회동	122.6957633	328

425 obs x 3 variables

이 데이터를 기반으로
클러스터링을 진행할 예정!

NEXT WEEK
To be Continued...





04

다음주 예고

우리의 주분은 끝나지 않았다.

다음주 예고



01.
주제 및
선정 배경

02.
전처리 및
EDA

03.
회귀 모델링

04.
다음주 예고

01

추가적인 변수 고려



02

다양한 클러스터링 모델링을 통해
우선 입지 행정동 선정



03

입지선정 알고리즘을 통한
우선 입지 선정

